IMPROVEMENT OF PROTEIN SEQUENCE ALIGNMENTS FOR BETTER PROTEIN STRUCTURE PREDICTION

Yen Hock Tan[1], Daisuke Kihara*[2, 1]
[1]Department of Computer Sciences, [2]Department of Biological Sciences
Purdue University, W. Lafayette, IN 47907
dkihara@purdue.edu

The quality of sequence alignments is a key for accurate protein structure prediction, because it is often the case that we need to align distantly related sequences. In fact, homology modeling methods heavily rely on the initial alignment, while in threading methods, errors in alignments of the query sequence and template proteins are often a major problem. Even in the majority of *ab initio* methods, an alignment procedure is an essential component, since they use local structure information obtained from a database search. To establish a robust and improved methodology of alignments, here we have examined several variants and parameters of dynamic programming algorithm. Calculated alignments are compared to a benchmark set of structurally aligned sequence sequences, prepared for three levels of the sequence homology, i.e., the family, the superfamily, and the fold level. First we investigated the accuracy of the optimal alignments using different distance matrices and gap penalties, because these are key factors which govern resulting alignments. We observed that the difference of the performance of the matrices become larger when the sequence similarity level becomes more distant. The effect of using multiple sequence alignments and sub-optimal alignments on the accuracy will be discussed.